

## COLLABORATIVE DEVELOPMENT IN WIKIPEDIA

Gerald C. Kane  
 Boston College  
 Chestnut Hill, Massachusetts  
 gerald.kane@bc.edu

Sam Ransbotham  
 Boston College  
 Chestnut Hill, Massachusetts  
 sam.ransbotham@bc.edu

**ABSTRACT**

Using 16,068 articles in Wikipedias Medicine Wikiproject, we study the relationship between collaboration and quality. We assess whether certain collaborative patterns are associated with information quality in terms of self-evaluated quality and article viewership. We find that the number of contributors has a curvilinear relationship to information quality, more contributors improving quality but only up to a certain point. Other articles that its collaborators work on also influences the quality of an information artifact, creating an interdependent network of artifacts and contributors. Finally, we see evidence of a recursive relationship between information quality and contributor activity, but that this recursive relationship attenuates over time.

**INTRODUCTION**

A new generation of IT-enabled collaborative tools — such as wikis, blog communities, and electronic social networks — enable people to share and create knowledge in new and potentially more powerful ways. These tools transcend traditional limitations and enable collective collaboration across conventional organizational boundaries. Nevertheless, the mere presence of these tools does not ensure effective collaboration or the creation of valuable knowledge. People and organizations must use these tools effectively to generate valuable outcomes.

Most of the extant research on these types of IT-enabled collaboration focuses on independent peer production communities working together to produce a single information artifact such as an open source software product or a single Wikipedia article. This research overlooks the fact that these collaborative environments often produce multiple information artifacts concurrently, and contributors may transfer knowledge from one artifact to another. Collaboration associated with the production of one information artifact may not be independent of the collaboration associated with other information artifacts on a shared collaboration platform. Understanding how collaboration occurs on one peer-produced information artifact can be important for understanding the quality of the other artifacts produced on the shared platform.

Wikipedia is becoming an increasingly important source of information for the general public, and it provides an excellent forum in which to examine collaborative practices (Kane & Fichman 2009). We examine the collaborative activities that occur between 16,068 articles and 40,479 contributors in the Wikipedia Medicine Wikiproject. We investigate whether the quality of an article is

associated with collaborative activity of its contributors to other Wikipedia articles. Furthermore, we examine whether there is a recursive relationship between information quality and the contributions an information artifact receives.

In general, we find support for our hypotheses. The collaborative processes that produce information artifacts on IT-enabled collaborative platforms are not independent from one another. Instead, their development is also influenced by the work of contributors on other information artifacts on the platform. Furthermore, we find a recursive relationship between information quality and collaboration. More contributors create better information artifacts which in turn attracts more contributors, but this relationship attenuates over time.

Future research examining the quality of peer produced information may be well-served by considering the interconnection between collaborative projects and the dynamics of that collaboration over time.

**THE ROLE OF INFORMATION ARTIFACTS**

Researchers have historically conceptualized IT-enabled collaborative environments as a network, examining how the structural features of those networks are associated with information benefits provided by that network (Ahuja, Galletta & Carley 2003, Wasko & Faraj 2005). An important difference between these previous collaborative environments and newer generations of IT-enabled collaborative tools is that people often use these emergent platforms to create peer-produced information artifacts, such as shared wiki articles, blog posts and comments, online videos and ratings, or interactive profile pages on electronic social network platforms.

These information artifacts preserve and extend the work of individual contributors and create a whole that is fundamentally different from the sum of its parts or the intentions of original contributors. A good example is how these information artifacts are often co-opted following a contributors death (Cohen 2009), when they cease being an outlet for the individual and become a place to commemorate and memorialize its original author. Thus, the peer-produced information artifacts may be used independently of the individuals who contributed to them or the purposes for which they were contributed. As such, these information artifacts may influence and be influenced by the contributors that work on them and thus can be considered as independent entities within the collaborative platform.

As contributors work on multiple information artifacts over time, they transfer information and knowledge contained in one artifact to another to improve the information quality of the recipient artifact. Three types of information and knowledge found in one artifact might be helpful to the development of another. Contributors can transfer content from one artifact to another that can be used to improve the information quality of the focal artifact in a number of ways. First, the content found in one information artifact may simply be used to improve the content of another. Second, contributors can also transfer process information from one artifact to another to improve the quality of the focal artifact. Based on their experience working on one information artifact, contributors may have learned effective ways to collaborate with others using the IT-enabled collaborative platform. For instance, if an individual has been involved in a large number of conflicts in other communities, s/he may have gained valuable insight on how to handle similar types of conflict in other communities. Finally, contributors can transfer reputational information about other contributors and how they contribute to (or detract from) the effective development of other information artifacts. If one contributor has a reputation for high-quality work, other contributors may be more willing to trust their insight than another contributor who is relatively unknown.

If information artifacts have an identity independent from the contributors who create it and if the contributors can transfer information gained from one artifact to another, the result is an interconnected network of contributors and information artifacts. The structure of this network might have significant implications for the quality of the information on the platform. Social Network Analysis (SNA) has been adopted by the organizational literature as a productive approach for studying these types of interconnected collaborative environments (Cross & Prusak 2002, Borgatti & Cross 2003, Reagans & McEvily 2003, Cummings 2004, Ransbotham, Kane & Lurie 2012).

### **Network Structure and Information Quality**

Three aspects of the collaboration that occurs in IT-enabled collaborative environments have been theorized as related to the quality of information it produces (Constant, Sproull & Kiesler 1996).

First, the number of contributors is often associated with the quality of information. Attracting a sufficient number of contributors is important for collaborative user-generated content. More contributors increase the effort and energy dedicated to creating content and provides a broader array of knowledge and abilities for content creation. This should increase the value of collaborative user-generated content. Research on prediction markets, virtual teams, and social networks suggests that the quality of aggregate information, number of ideas generated, and likelihood of a valuable answer increases with

the number of participants (Constant et al. 1996, Martins, Gilson & Maynard 2004, Foutz & Jank 2010).

At the same time, other research suggests that having too many contributors can also be problematic. After a certain point, the marginal cost of adding new members exceeds its marginal value. Consistent with the adage “too many cooks spoil the stew,” an excessive number of contributors negatively influence the value of user-generated content. As the number of contributors grows, the marginal value of additional contributors decreases while the cognitive and coordination costs associated with contributions increases (Asvanund, Clay, Krishnan & Smith 2004, Jones, Ravid & Rafaeli 2004, Ransbotham & Kane 2011). In particular, those involved in the co-creation of content are likely to suffer from information overload as they try to make sense of and respond to others contributions.

This rationale suggests that a curvilinear relationship between number of contributors and information quality. The most valuable collaborative user-generated content is generated when enough contributors are attracted to sustain production but not so much that it creates information overload for contributors. Considerable empirical evidence supports such curvilinear relationships between number of contributors and outcomes in online collaborative groups (Butler 2001, Hansen & Haas 2001, Asvanund et al. 2004, Oh & Jeon 2007, Ransbotham & Kane 2011). Similar relationships have also been found in traditional organizations. For instance, software development teams often need sufficient resources to accomplish their goals, but adding more members to a troubled or delayed project can compound delays by increasing coordination costs (Brooks 1975), often exponentially, as new members are added (Espinosa, Slaughter, Kraut & Herbsleb 2007). Thus, we expect a curvilinear relationship between the number of contributors and the quality of collaborative user-generated content. These ideas lead to our first hypothesis:

**HYPOTHESIS 1.** *The quality of a peer-produced information artifact will be curvilinearly (inverted-U) related to the number of contributors to the artifact.*

Second, greater diversity of information sources provided by an IT-enabled collaborative environment improves the quality of information it produces (Constant et al. 1996). Additional contributors may not be particularly valuable if they provide the same information already possessed by other contributors, but they are valuable when they provide access to information not already possessed by existing contributors (Burt 1997, Uzzi 1997, Kane & Alavi 2007).

The number of different information artifacts that contributors work on reflects the diversity of information available to the information artifact on a mass collaboration. It can reflect both the knowledge directly available for transfer from other information artifacts or it may simply serve to reflect the underlying knowledge

possessed by the individual contributors. The diversity of knowledge possessed by individual contributors on an IT-enabled collaborative platform may be revealed in the pattern of other artifacts the contributors work on. For instance, contributors with deep, specialized knowledge may work intensely in a few communities with related purposes; whereas contributors with broad, more generalized knowledge may work more superficially on a broad range of other artifacts.

Further, these patterns reveal the type of knowledge possessed by the contributor. For example, in the wake of the Virginia Tech Massacre, contributors reported very different reasons for contributing to the related Wikipedia article (Kane & Fichman 2009). Some contributors did so because of their knowledge of the school, some because they had knowledge and interest regarding the relevant gun control issues, and still others because they were skilled copyeditors. The first type of contributor may also contribute to articles on Virginia or other colleges, the second type might also contribute to other gun-related topics, and the third type may contribute to a diverse range of articles of a particular length or stage of development. Thus, the other information artifacts on which a contributor works on reflects the underlying knowledge and/or topical interests possessed by that contributor. We hypothesize that the number of different information artifacts on which a contributor in a peer-production community also works reflects the diversity of information sources available to the community.

**HYPOTHESIS 2.** *The quality of a peer-produced information artifact will be positively related to the number of other information artifacts on which its contributors work.*

The depth of resources available in a collaborative environment will also be related to the quality of information it produces (Constant et al. 1996). Even if a collaborative environment provides access to a large number and to a diverse range of information sources, some sources have deeper and more valuable resources than others. Certain contributors simply provide access to greater information resources, either as a result of their connection to information artifacts with more resources or as a result of the underlying resources possessed by that individual. Access to deeper resources generated in more active peer-production environments positively relates to the quality of information produced by the community. In many IT-enabled collaborative platforms, a relatively small percentage of information artifacts accounts for a relatively large amount of the collaborative activity that occurs on it (Kuk 2006). Information artifacts that are the source of more collaborative activity are deeper sources of valuable content, process, and reputational information.

Similarly, individuals who are more influential contributors to artifacts that host abundant collaborative activity may also reveal the underlying depth of resources possessed by the individual. IT-enabled collaborative platforms typically employ limited hierarchical and ad-

ministrative structures, if they possess any at all (Butler, Joyce & Pike 2008). Individuals who emerge as prominent contributors do so largely because members of the community recognize them as valuable contributors. In peer-produced information artifacts, someone can contribute heavily only by the consent of other contributors. If other contributors do not approve of an individual's contributions, they will either resist them, forcing the unwelcome contributor to relent or leave (Kane, Majchrzak, Johnson & Chen 2009); or else the other contributors will leave, as they are no longer receiving benefits from the collaborative community (Butler 2001).

Thus, within a IT-enabled collaborative platform, the activity level of contributors on prominent information artifacts reflects the depth of resources available to a peer production community. More active information communities are the source of deeper content, process, and reputational information than less active communities. Peer-produced artifacts with access to deeper resources are more likely to produce higher quality information.

**HYPOTHESIS 3.** *Depth of collaborative activity that occurs within an information artifacts collaborative network will be positively related to artifact quality.*

While we hypothesize above that collaboration leads to improved quality of the information artifact, it is also possible that the quality of the information artifact will also lead to certain types of collaborative behavior (Kane et al. 2009). As articles become of higher quality, they are more likely to attract interest from outsiders who seek to access that information, either for the content or as a collaborative exemplar. In open collaborative environments, all viewers of an article are also potential contributors. Higher quality information may, therefore, attract more viewers, introducing the possibility that these viewers in turn become collaborators. Thus, while we have hypothesized that certain collaborative structures lead to improved information quality, it is also possible that improved information quality will in turn lead to certain collaborative patterns. Thus, we hypothesize a recursive relationship between the quality of an article and the collaboration it generates.

**HYPOTHESIS 4.** *There will be a recursive, positive relationship between information quality and the collaboration that occurs on an information artifact.*

## **RESEARCH METHOD AND SETTING**

To test our hypotheses, we employ social network analysis (SNA). Social network analysis is capable of examining more complex networks comprising different types of nodes (Wasserman & Faust 1994). A traditional but infrequently-used network conceptualization is known as the two-mode network (Borgatti & Everett 1997, Faust 1997). A two-mode network is a general network structure consisting of two fundamentally different types of entities that cannot be examined equivalently with one another. Here, we conceptualize our two-mode network as consisting of the information

artifacts and individual collaborators as nodes; editing activities are the ties that connect them, due to the transfer information and knowledge from one artifact to another.

We use two different network measures to operationalize our two remaining hypotheses — degree centrality and eigenvector centrality. These centrality measures are often used in conjunction to capture the features of the local (degree) and the global (eigenvector) social network (Friedkin 1991, Faust 1997). Degree centrality in the two-mode network is used to measure the diversity of information sources available to an information artifact through its contributors. Eigenvector centrality captures the depth of information sources available to an information artifact. This measure summarizes the node's centrality in the global network of all of the nodes and ties that compose the network.

We focus empirical analysis on the 16,068 articles within the Medicine Wikiproject in Wikipedia. A Wikiproject is a group of contributors dedicated to develop, maintain, and organize articles related to a particular topic. We focus on a single Wikiproject, because a random sample of articles would not likely yield the social network features of theoretical interest and a Wikiproject provides clearly defined boundaries.

### Data Collection

We downloaded the full text history of 2,029,443 revisions of 16,068 articles by 40,479 unique contributors in the Medicine Wikiproject as of June 2009, which resulted in a 50 GB data set of raw data. We employed a 70-node Linux cluster to allow for simultaneous downloads and processing of these extensive data. For each contribution, we record the contributors identity, the changes made, a description of the change, and the time of the change.

### Dependent Variables

We use two different measures to evaluate the quality of an article. First, we assess self-evaluated quality. The Medicine Wikiproject evaluates articles on a 7-point scale (Stub, Start, C, B, Good, A, Featured). We recruited two fourth year medical school students to independently validate the quality of a subsample of 120 randomly selected articles. Each student independently evaluated each article, then ratings were compared and reconciled to create a single reviewer rating. These reconciled ratings were then compared to the ratings assigned by the Wikiproject Medicine. These students reached an 85% interrater reliability with one another, and the reconciled ratings achieved a 90% agreement with the ratings of information quality assigned by the Wikiproject. These results suggest the self-evaluated quality was a good proxy for the overall quality of the article.

Second, we also use public-evaluated quality in order to provide an independent and finer-grained measure of

Table 1. Self-Evaluated Quality

Variable	Model 0	Model 1
Importance (medium)	0.532*** (0.044)	0.537*** (0.044)
Importance (high)	0.785*** (0.080)	0.736*** (0.082)
Importance (top)	1.383*** (0.152)	1.251*** (0.161)
Age (ln, years)	0.137*** (0.036)	0.103** (0.036)
Daily Views (ln, views)	0.218*** (0.033)	0.180*** (0.033)
Length (ln, characters)	1.876*** (0.094)	1.753*** (0.097)
Complexity (ARI)	-4.191*** (0.488)	-4.366*** (0.493)
Section Depth	0.144*** (0.028)	0.135*** (0.028)
External References	0.253*** (0.032)	0.221*** (0.030)
Multimedia content	-0.063 (0.048)	-0.067 (0.051)
Anonymity (percentage)	0.905*** (0.149)	1.029** (0.149)
Distinct Contributors		0.269*** (0.044)
Distinct Contributors (squared)		-0.186*** (0.030)
Degree Centrality per Contributor		0.168*** (0.034)
Eigenvector Centrality		0.105*** (0.022)
Log pseudolikelihood	-9,036.96	-8,949.33
Wald $\chi^2$	4,294.34***	4,332.24***
Pseudo $R^2$	36.74	37.30

quality that provided richer data for testing Hypothesis 4. Here, we operationalized information quality as the number of times an article has been viewed, which is an indication of how this information is valued by the public (Ransbotham et al. 2012). For each article, we collected the number of views each day from December 2007 until June 2009; these data are not available for the entire history of Wikipedia. We summarized the view counts by month.

### ANALYSIS AND RESULTS

The full dataset was then analyzed using Stata. Ordinal regression is appropriate when there is a progressive relationship within a categorical dependent variable, but it is unclear the magnitude of different between the categories. For instance, the observer may know which Olympic athletes have won the gold, silver, and bronze medals without knowing the final scores of any of the athletes. This method is most appropriate for our measure of self-evaluated quality. Table 1 describes the full results of an ordinal logistic regression on self-evaluated

Table 2. Three Stage Least Squares Model of Article Views

Model	Variable	Model 1		Model 2	
Equation 1: Article Views (ln/1000)					
	Monthly Fixed Effects	indicators		indicators	
	Constant	744.439***	(5.215)	755.126***	(5.387)
	Article Views (ln, lagged)	0.971***	(0.001)	0.971***	(0.001)
	Age (ln, years)	18.287***	(1.285)	17.964***	(1.384)
	Length (ln, characters)	9.135***	(1.164)	8.904***	(1.166)
	Complexity (ARI)	-27.702*	(11.181)	-26.060*	(11.181)
	Section Depth	1.975*	(0.847)	1.795*	(0.848)
	External References	1.038	(0.690)	0.629	(0.693)
	Internal Links	3.261**	(1.129)	2.990**	(1.135)
	Multimedia Content	-0.014	(0.661)	-0.031	(0.661)
	Anonymity (percentage)	91.676***	(5.857)	75.019***	(6.171)
	Contributors	12.280***	(1.424)	48.579***	(4.627)
	Contributors <sup>2</sup>	-6.780***	(1.115)	-61.869***	(6.901)
	Local Centrality	-1.993**	(0.672)	-2.738***	(0.729)
	Global Centrality	3.266***	(0.685)	1.908*	(0.765)
	Age × Contributors			-27.487***	(3.483)
	Age × Contributors <sup>2</sup>			42.469***	(5.158)
	Age × Local Centrality			-0.856	(0.747)
	Age × Global Centrality			-1.694***	(0.391)
	R <sup>2</sup>	98.099		98.100	
	χ <sup>2</sup> (×10 <sup>6</sup> )	6.44***		6.44***	
Equation 2: Contributors					
	Monthly Fixed Effects	indicators		indicators	
	Constant	-0.250***	(0.053)	-0.251***	(0.053)
	Contributors (lagged)	1.021***	(0.001)	1.021***	(0.001)
	Article Views (ln)	0.103***	(0.006)	0.103***	(0.006)
	Age (ln, years)	-0.121***	(0.011)	-0.121***	(0.011)
	Length (ln, characters)	0.002	(0.012)	0.002	(0.012)
	Complexity (ARI)	-1.034***	(0.113)	-1.034***	(0.113)
	Section Depth	0.041***	(0.009)	0.041***	(0.009)
	External References	-0.052***	(0.007)	-0.052***	(0.007)
	Internal Links	0.139***	(0.011)	0.139***	(0.011)
	Multimedia Content	-0.025**	(0.007)	-0.025**	(0.007)
	Anonymity (percentage)	0.407***	(0.056)	0.407***	(0.056)
	Article Protected? (1=yes)	-5.787***	(0.087)	-5.789***	(0.087)
	R <sup>2</sup>	99.958		99.958	
	χ <sup>2</sup> (×10 <sup>8</sup> )	3.02***		3.02***	

124,711 observations; standard errors in parentheses; significance \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

quality. Model 0 presents our results for the baseline model with only control variables, and Model 1 presents the results of our models with the variables of interest. (We also tested each independent variable of interest independently, and results are consistent with the composite findings in Model 1.)

Examining Hypothesis 1, both the linear and square coefficients are significant ( $\beta = 0.269$ ,  $p < 0.001$  and  $\beta = -0.186$ ,  $p < 0.001$ , respectively). These coefficients indicate an inverted-U relationship with article quality. Additional contributors working on an information artifact increase quality up to an optimal point, but then additional contributors detract from the quality of information found in the artifact. We also find support for the second hypothesis that diversity of resources increases artifact quality. The coefficient on degree centrality per contributor is positive and significant ( $\beta = 0.168$ ,  $p < 0.001$ ). The more diverse the content, process, and reputational knowledge contributors access and/or represent, the higher quality the information artifact is likely to be. Hypothesis 3 that depth of resources available to an information artifact will be positively related to information quality is also supported. The coefficient on eigenvector centrality is positive and significant ( $\beta = 0.105$ ,  $p < 0.001$ ). The greater depth of resources that an artifact can access through or is represented by its contributors, the higher quality the artifact is likely to be.

Table 2 describes the results of a simultaneous equation, three stage least squares regression on the natural log of article views (scaled by 1,000 for presentation) using the sample of 124,711 monthly observations of articles from December 2007 until June 2009. Model 1 introduces the focal network variables. Because of the large sample size, we use a low threshold of statistical significance ( $p < 0.001$ ) to test our hypotheses.

We find partial support for Hypothesis 4. The number of unique contributors to user-generated content has a curvilinear relationship with article views. Both the linear and squared coefficients are significant ( $\beta = 12.28$ ,  $p < 0.001$  and  $\beta = -6.78$ ,  $p < 0.001$ , respectively). These coefficients indicate an inverted-U relationship with article views. Additional contributors working on an article increase its quality up to an optimal point, but then detract from the ability of the article to attract viewers. We also considered models that utilized either a linear effect of contributors or a log of the number of contributors; based on the Akaike information criterion (AIC); the quadratic model provides a slightly better fit (decreases in AIC of 18 and 7 respectively).

However, we do not find that the quality of user-generated content is positively related to local network centrality; the coefficient for degree centrality per contributor is negative and significant ( $\beta = -1.99$ ,  $p < 0.010$ ). We do find that the quality of user-generated content is positively related to global network centrality; the coefficient for eigenvector centrality is positive and

significant ( $\beta = 3.27$ ,  $p < 0.001$ ). As hypothesized, both models demonstrate a recursive effect of article viewership on the number of contributors. In Equation 2, the coefficient for article views is significant and positive ( $\beta = 0.10$ ,  $p < 0.001$ ). More contributors lead to greater viewing, but more viewing also yields a greater number of contributors. The protect variable, used for identification of the simultaneous model since it affects contributions but not viewing, is also significant ( $\beta = -5.79$ ,  $p < 0.001$ ) and behaves as expected. When an article has restrictions placed on who can contribute, significantly fewer people contribute to it.

It is interesting to note that age has the opposite effect in the contributor model — age is positively related to viewing but negatively related to the overall number of contributors. This suggests that collaborative user-generated content matures and stabilizes over time; more people come to view older content but they are less likely to contribute to that content. It may be that more mature content attracts a more general audience that is less likely to have the knowledge or inclination to contribute, or it may be that the viewers of the content find it to be relatively complete and feel they have nothing to add to improve it.

## CONCLUSION

In this paper, we test the influence of the network structure created by contributors and information artifacts on information quality in peer-produced information. We find good general support for our hypotheses, determining that the collaboration occurring on one information artifact can influence the quality of the other information artifacts on which those collaborators work. Furthermore, we also demonstrate a recursive relationship between contributors and information quality in terms of viewership. More viewers brings more collaborators which brings more collaborators, but this recursive relationship attenuates over time. Implications are that researchers should broaden their understanding of how collaboration on other information artifacts can influence information quality and begin understanding peer-production settings as a network of knowledge.

## ACKNOWLEDGMENTS

Gerald Kane acknowledges funding for this research from the National Science Foundation (CAREER 0953285).

## REFERENCES

- Ahuja, M., Galletta, D. & Carley, K. (2003), 'Individual Centrality and Performance in Virtual R&D groups: An Empirical Study', *Management Science* 49(1), 21–38.
- Asvanund, A., Clay, K., Krishnan, R. & Smith, M. (2004), 'An Empirical Analysis of Network Externalities in Peer-To-Peer Music-Sharing Networks', *Information Systems Research* 15(2), 155–174.

- Borgatti, S. P. & Cross, R. (2003), 'A Relational View of Information Seeking and Learning in Social Networks', *Management Science* **49**(4), 432–445.
- Borgatti, S. P. & Everett, M. G. (1997), 'Network Analysis of 2-mode Data', *Social Networks* **19**(3), 243–269.
- Brooks, F. P. (1975), *The Mythical Man-Month: Essays on Software Engineering*, Addison-Wesley, Reading, MA.
- Burt, R. (1997), 'The Contingent Value of Social Capital', *Administrative Science Quarterly* **42**(2), 339–365.
- Butler, B., Joyce, E. & Pike, J. (2008), Don't Look Now, But We've Created a Bureaucracy: The Nature and Roles Of Policies and Rules in Wikipedia, in 'Proceeding of the Twenty-Sixth Annual SIGCHI conference on Human Factors in Computing Systems', ACM, pp. 1101–1110.
- Butler, B. S. (2001), 'Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures', *Information Systems Research* **12**(4), 346–362.
- Cohen, A. (2009), 'A New Kind of Memorial for the Internet Age', *The New York Times* . 25 July 2009.
- Constant, D., Sproull, L. & Kiesler, S. (1996), 'The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice', *Organization Science* **7**(2), 119–135.
- Cross, R. & Prusak, L. (2002), 'The People Who Make Organizations Go— or Stop', *Harvard Business Review* **80**(6), 104–112.
- Cummings, J. N. (2004), 'Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization', *Management Science* **50**(3), 352–364.
- Espinosa, J., Slaughter, S., Kraut, R. & Herbsleb, J. (2007), 'Team Knowledge and Coordination in Geographically Distributed Software Development', *Journal of Management Information Systems* **24**(1), 135–169.
- Faust, K. (1997), 'Centrality in Affiliation Networks', *Social Networks* **19**(2), 157–191.
- Foutz, N. Z. & Jank, W. (2010), 'Prerelease Demand Forecasting for Motion Pictures using Functional Shape Analysis of Virtual Stock Markets', *Marketing Science* **29**(3), 568–579.
- Friedkin, N. (1991), 'Theoretical Foundations for Centrality Measures', *American Journal of Sociology* **6**, 1478–1504.
- Hansen, M. T. & Haas, M. R. (2001), 'Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company', *Administrative Science Quarterly* **46**(1), 1–28.
- Jones, Q., Ravid, G. & Rafaeli, S. (2004), 'Information Overload and The Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration', *Information Systems Research* **15**(2), 194–210.
- Kane, G. & Alavi, M. (2007), 'Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes', *Organization Science* **18**(5), 796–812.
- Kane, G. C. & Fichman, R. G. (2009), 'The Shoemaker's Children: Using Wikis for Information Systems Teaching', *MIS Quarterly* **33**(1), 1–17.
- Kane, G. C., Majchrzak, A., Johnson, J. & Chen, G. (2009), A Longitudinal Study of Perspective Development in a Fluid Online Collective, in 'Proceedings of the 30th International Conference on Information Systems', Phoenix, AZ.
- Kuk, G. (2006), 'Strategic Interaction and Knowledge Sharing in the KDE Developer Mailing List', *Management Science* **52**(7), 1031–1042.
- Martins, L. L., Gilson, L. L. & Maynard, M. T. (2004), 'Virtual Teams: What Do We Know and Where Do We Go From Here?', *Journal of Management* **30**(6), 805–835.
- Oh, W. & Jeon, S. (2007), 'Membership Herding and Network Stability in The Open Source Community: The Ising Perspective', *Management Science* **53**(7), 1086–1101.
- Ransbotham, S. & Kane, G. C. (2011), 'Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia', *MIS Quarterly* **35**(3), 613–627.
- Ransbotham, S., Kane, G. C. & Lurie, N. H. (2012), 'Network Characteristics and the Value of Collaborative User-Generated Content', *Marketing Science* **31**(3).
- Reagans, R. & McEvily, B. (2003), 'Network Structure and Knowledge Transfer: The Effects of Cohesion and Range', *Administrative Science Quarterly* **48**(2), 240–267.
- Uzzi, B. (1997), 'Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness', *Administrative Science Quarterly* **42**(1), 35–67.
- Wasko, M. & Faraj, S. (2005), 'Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice', *MIS Quarterly* **29**(1), 35–57.
- Wasserman, S. & Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge, New York.